ATTACHMENT U) 2 PAGES

TECHNOLOGY SECTION 2
ELECTRONIC ENGINEERING TIMES

Though It's Still A Curiosity, Hypercube's Making Progress

By Ray Weiss

PASADENA, Calif. — Following the carnage of the Rose Bowl, Pasadena last week played host to a different sort of gathering, the third annual Hypercube Conference. There, vendors, software developers, researchers and users gathered to thresh out the problems of the emerging architecture.

This year, three of the major vendors, Floating Point Systems Inc., Intel Scientific Computing and Ncube Corp. (all of Beaverton, Ore.) also were exhibitors. A fourth, Ametek Computer Research Division (Monrovia, Calif.), held local off-conference presentations.

Hypercubes are relatively new, with the first commercial system, Intel's iPSC, released in 1985. Yet with over 250 papers to be given or distributed and over 700 attendees, the conference shows the response of the research community.

Just Getting Started

Andrew Witkowski of the Jet Propulsion Laboratory, the coordinator of the conference, believes that the wide range of papers shows that the use of the hypercube is just getting started.

"Hypercube is the most successful distributed architecture, with far more expandability than the parallel shared-memory systems," he says. It just takes time for the necessary software and applications to develop.

Hypercubes, says Witkowski, are now being applied to scientific and engineering problems. This is the traditional path that new architectures take, including the early IBM machines. These arenas are willing to take on the burden of new machines, even with limited software, if they supply cost-effective processing power. The next use of hypercubes, he states, will be in more mainline applications like database and transaction processing.

Papers at the conference explore a number of extensions to the hypercube including real-time operating systems, distributed Prolog, parallel AI processing and simulation. One, for example, presents Interwork II, a software package that sits above the Intel node operating system that creates a global object space across the hypercube. It also provides lightweight processes (fast, low-overhead tasks and global synchronization for simulation). The software can be run on PCs for development and later transferred to a hypercube for production runs.

Hypercube is a multiprocessor architecture that was born at the California Institute of Technology and fathered by research teams headed by Charles Seitz. It relies on loosely coupled processors. Each node processor has its own memory and is linked into a binary n-cube configuration capable of supporting up to 1,024 nodes. Coordination between nodes is done via message passing. Earlier hypercubes relied on passing messages down a nearest-neighbor chain rather than point-to-point communications.

Existing hypercube implementations tend to use off-the-shelf microprocessors, usually in single-board configurations, as hypercube nodes. An exception is the innovative Ncube architecture that uses a proprietary 32-bit processor and communications chip as a node, combining seven- or 10-chip nodes on a single board. However, this architecture limits the node memory to 512 kbytes.

Solving Access Problems

The current hypercube architectures are concentrating on solving access problems to the individual processor nodes. Both Intel and Ametek have improved node communications by going to different point-to point-routing schemes that rely on separate node-communication controllers to speed up communications and off-load the node processor.

Both Ncube and Ametek are providing solutions that enable individual nodes or clusters of nodes to do direct-disk I/O, bypassing the host-system controller. "This approach," explains Ncube president John Palmer, "allows the nodes to stage data for processing using multiple disk drives. It also opens up the hypercube for more traditional processing that requires local disk storage."

Ncube is adding a parallel-I/O system, the Ncube NChannel, an I/O board that supports 16 serial I/O channels. Each is capable of driving a 4-Mbyte/s controller, and each channel is controlled by an Ncube chip. The same chip is used in the Ncube as a node processor. Because of its high integration, the Ncube is the first of the commercial hypercubes to be configured with as many as 1,024 nodes, which can deliver up to 2,000 Mips.

At the conference, Ncube announced a major price reduction for its low-end four-node card for the PC. The board will be reduced to $9,995, from $20,000, to introduce engineers to hypercube processing. For the universities, the board will be further discounted to $4,995 to encour-

(Continued on Page 42)

Ametek's Backplane As Hypercube Rival?

Ametek, backed up with research by Cal Tech's Charles Seitz, brought out a design that can support up to 1,024 nodes on a mesh, an X-Y coordinate backplane with each node plugging into a special communications chip. Accordingly, a high-speed X-Y backplane could be as effective as a hypercube, if the backplane bandwidth is high enough, says Jeff Fier, Ametek applications manager.

The 2010 relies on a special chip, the AMRD, which provides a four-channel (north, south, east, west) routing path for each node. It can support up to four concurrent channels, each running at 20 Mbytes/s. Routing in the box is done by hardware implementing a simple X-Y routing protocol, which can delay the 256-byte packets up to 2 μs at each node. Thus maximum end-to-end transmission for a 1,024-node system is 128 μs.

"We have removed the limitations inherent in the first- and second-generation hypercube systems," explains Yin Shih, president of the Computer Research Division. The approach allows point-to-point communications between nodes, while eliminating the hypercube interconnections that required node growth to be a power of two. Now, notes Shih, it is easy to incrementally add nodes.

The original Ametek box, followed the pioneering Cal Tech design, using Intel 286 processors for each node. The new design uses Motorola 25-MHz 68020s, with a custom MMU, a 68881 or 68882 FPU and up to 8 Mbytes of RAM. An optional vector floating-point processor is available (20 Mflops peak single- and double-precision) for each node. Additionally, each node can accommodate up to three VME boards, allowing disks to be attached to individual nodes. The design will be in the hands of beta sites in the third quarter, with production deliveries in October.

Superconductivity In Search Of Reasons Why

(Continued from Page 41)

Robert Schrieffer of the University of California at Santa Barbara is continually renewing BCS to keep up with research findings. "I think of the BCS theory as the general framework for pairing," Schrieffer maintained. There are many possible applications of it to specific models. "For 25 years, John Bardeen, Leon Cooper and I have tried to make that point." The pairing effect needs only a force to make it effective. In different mediums, different forces can cause the effect. The force that explains high-temperature superconductivity is different from the low-temperature superconductivity which Bardeen, Cooper and Schrieffer explained.

BCS originally posited that at very low temperatures (4°K), particles that normally repel will be drawn into pairs, but not groups of three or larger. At higher temperatures, thermal fluctuations keep these particles apart. But later it was seen that at higher temperatures (to 35°K), a second force of attraction exists. When two particles are drawn together, the total energy is lowered. One might think that lowering the temperature would attract three or four together, "but it turns out to be either two or infinity," Schrieffer explained. And when they condense, they condense to infinity and into superconductivity.

The first application of BCS was to the low-temperature superconductivity of metallic semiconductors. Here, it was applied to the "elastic" force of phonon attraction. The concept is similar to the way two metal balls will make two small round indentations ("holes") in an elastic sheet and then roll together into a single deeper hole. Similarly, electron-holes (phonons) are attracted under suitable conditions by the same elastic force. If the temperature is low enough, then the elastic attraction brings pairs of phonons together and as they condense the metal becomes superconducting.

In semiconductors at higher temperature, the "elastic sheet" is replaced by an electron sea that exists between the layers of the material used. The layers are usually arranged like a checkerboard of copper molecules, which act as antiferromagnets. The polarity of the magnets alternates from molecule to molecule, and oxygen molecules separate the copper. The entire layer is doped with impurities to create holes.

Properly spaced antiferro magnets, oscillating from plus to minus, make that layer into an insulator. "That has been known for 40 years. What is new is that if you dope that material to add extra electrons or withdraw electrons, putting in extra holes, these added carriers, electrons or holes, are attracted to each other by a magnetic analog of the elastic rubber sheet that causes low-temperature metals to superconduct," Schrieffer commented.

The key to the arrangement is that the distance between the copper molecules doesn't change as holes are added (electrons are taken out), allowing pairs of electrons through. As single particles, the electrons are trapped between the layers, where their charge becomes separated from their spin. (Spin, like charge, is a property of an electron, and indicates the direction in which a particle rotates. Stationary electrons have "balanced" spins, rather than no spin.) So-called spin bags encapsulate the charge and spin of a trapped particle. The surprise is that not only electrons but also electron-holes can be "clothed" in a spin bag. "When you pull an electron out to form a hole, you also pull out spin and you leave spin behind," Schrieffer maintained.

If you pull one of those pairs—a single electron with spin—you leave the other half of spin pair behind, whose direction is opposite that of the electron's spin. Thus, created holes have a spin that is opposite that of the electron that departed to make the hole. These electron holes are what hops from plane to plane while its spin bag is left behind on one side and reforms on the other.

Anderson's theory holds that electrons totally separate their charge from their spin. Some have spin but no charge and others have charge but no spin. Both are solitons, particle-like excitations that move around between layers. When the sea of electrons between the layers shifts (it does so regularly) and leaves a hole of charge without any spin, that is a holon. When the sea moves back and leaves a little hole of spin without any charge, that is a spinon. Those are the two types of solitons in Anderson's theory.

Solitons are not free particles, except within the layers. planes. "Solitons are particularly strange in that they only live in the layers, you cannot take them out individually into the outside world because they disappear." But they can come out in pairs.

When holons pair up, the material goes superconducting. "The amazing thing is that the mathematics of this is the same as the mathematics of quarks. Quarks are solitons in a vacuum." Paired holons tunnel between planes, thus causing superconductivity. Each member of a pair moves independently, one following the other, in a coherent stream without losing relative phase.

Holons cannot tunnel singly without leaving behind an imprint of spin, which would expend a great deal of energy (resistance). But by pairing with another holon that jumps immediately after the first, the imprint is canceled by its opposite, with zero total energy expended (resistanceless).

But while the paired holons leave no trace behind, they upset things in the layer to which they jump. When solitons appear on the other side of the layer they are spinless and as such cannot reform into two normal electrons.

How do they regain their missing spin? To compensate for their zero spin locally, a slight anti-parallel spin is induced on the background. If the electron's spin is "up," then the background particles are adjusted "down."

Soliton advocates are only concerned about the local neighborhood of compensation, and the missing spin is thought of as being off elsewhere at a distance.

Hypercube Making Headway

(Continued from Page 41)

age research.

Intel, the prime supplier of hypercube systems with over 80 installations, will be showing its second-generation iPSC-2. Announced last August, it is now shipping. It was converted from a 16-bit 286 to a 32-bit 386 with a 387 coprocessor and up to 16 Mbytes per node, as well as an optional vector processor. With 128 vector-processor-equipped nodes, Intel specifies a max performance of 256 Mips with 1,280 single- and 427 double-precision megaflops.

Significantly improving performance was a new point-to-point communications scheme that doesn't use node-processing resources. Instead, a circuit switch establishes a path through the hypercube; the direct connection passes a serial stream point to point. Each node has a direct-connect module that handles up to 5 Mbytes/s of communications bandwidth.

Floating Point Systems showed its Transputer-based T Series machine. By midyear it will be shipping the Mark 2, which uses the Inmos T800 Transputer (running at 20 MHz) and the Weitek vector processor. For 128 nodes, the machine is expected to deliver performance topping 2 Gflops, according to Tom Bauer, T Series business manager. The system can be expanded up to 4,096 nodes.

The T Series takes advantage of the Transputers communications links, but lines must be multiplexed for the larger configurations. Overhead due to communications is limited to between 10 percent and 20 percent per node.

Not all analysts believe that hypercubes will have a smooth transition toward general applications. Omri Serlin, editor of the *Serlin Report on Parallel Processing*, believes that hypercubes face a major stumbling block in that they cannot run the existing engineering and scientific applications. "The so-called 'dusty decks' representing years of finished applications can't be run as is, they must be recoded, which give a tremendous advantage to alternative, closed-box processing solutions like those from Convex, Sequent, Multiflow and Alliant."

Until there are turnkey applications available for the hypercube, he says, it will be rough going for them. Currently, the vendors, especially Intel, are working to attract engineering software companies to port their turnkey application packages to a hypercube. Intel has gathered applications like the Nektron and Passage computational fluid dynamic packages.

Nevertheless, he maintains that hypercube is an interesting technology, capable of packing expandable processing power into a small package. He expects that in the long run, as software appears for the systems, they will become a viable alternative to large-scale processing. "Hypercubes have a major advantage in their loosely coupled architectures in that they are easily expandable and at the same time present a minimal path between any two nodes."